

Rasch analysis of the Western Ontario MacMaster Questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia

Frederick Wolfe, Sheldon X Kong

Abstract

Objective—Advances in health measurement have led to the application of Rasch Item Response Theory (IRT) analysis (Rasch analysis) to evaluate instruments measuring health status and quality of life of patients, including the Health Assessment Questionnaire and SF-36. This study investigated the extent to which the Western Ontario MacMaster osteoarthritis questionnaire (WOMAC) satisfies the Rasch model, particularly in respect to unidimensionality, item separation, and linearity.

Methods—The study included a total of 2205 patients, 1013 with rheumatoid arthritis (RA), 655 with osteoarthritis of the knee or hip (OA), and 537 with fibromyalgia. All patients completed the WOMAC as part of a longitudinal study of rheumatic disease outcomes. To examine whether the WOMAC pain and function scales each fits the Rasch model, the Winsteps program was used to assess item difficulty, scale unidimensionality, item separation, and linearity.

Results—Although the WOMAC worked best in OA, regardless of disorder, both the pain and function scales were unidimensional, had adequate item separation, and had a long range (25–150) of linearity in the function scale. Several functional items, however, had a high information weight fit (INFIT) statistic, indicating

poor fit to the model. These items included “getting in and out of the bath” and “going down stairs.”

Conclusion—The WOMAC generally satisfies the requirements of Rasch item response theory across all disorders studied, and is an appropriate measure of lower body function in OA, RA and fibromyalgia. Although some individual items do not fit well, it is not likely that removing such items would result in more than overall minimal differences, and it will be difficult to remove traces of multidimensionality while keeping the central constructs of progressive lower body musculoskeletal abnormality intact. In addition, it is possible that a “purer”, still more unidimensional instrument would be less useful in clinical trials and epidemiological studies by restricting the range of the scale.

(Ann Rheum Dis 1999;58:563-568)

Arthritis Research Center and University of Kansas School of Medicine, Wichita, Kansas, USA
F Wolfe

Merck and Co Inc, Whitehouse Station, New Jersey, USA
S X Kong

Correspondence to:
Dr F Wolfe, Arthritis Research Center, 1035 N Emporia, Suite 230, Wichita, KS 67214, USA.

Accepted for publication
26 April 1999

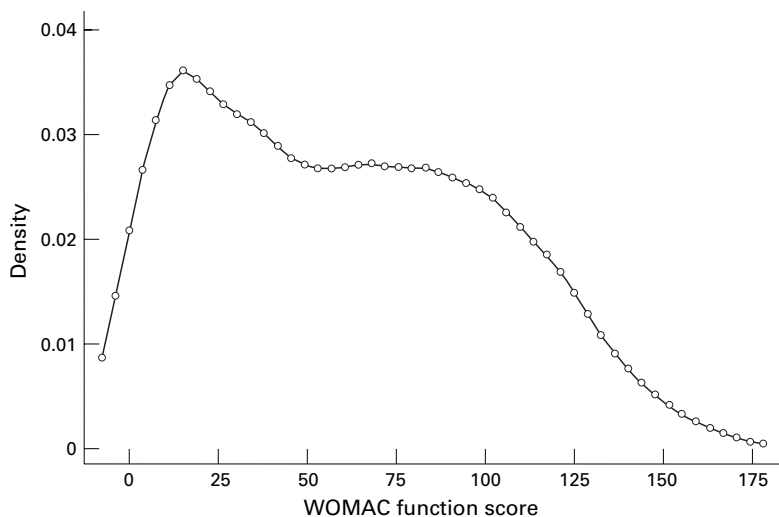


Figure 1 Distribution of 2205 WOMAC function scores as modelled in a kernel density plot. Actual range is 0–170.

Among the major determinants of health related quality of life in osteoarthritis of the knee or hip (OA) are pain and functional loss. The Western Ontario MacMaster (WOMAC) questionnaire was designed to measure these components by assessing 17 functional activities, five pain related activities, and two stiffness items.¹ This instrument has been widely studied, and many of its psychometric properties are known.²⁻⁶ It has been widely used in clinical trials^{7,8} because of its sensitivity to change and its construct validity.

Previous studies of function in OA have shown that WOMAC is more sensitive to change and has greater efficiency than most other instruments used to assess OA, including the Health Assessment Questionnaire (HAQ) and the SF-36 Health Survey.^{1,8-10} This is not surprising as the HAQ was developed and validated in a rheumatoid arthritis (RA) population, and the SF-36 was developed as a generic instrument for use in all populations. Despite this the WOMAC, like other instruments, does not correlate well with radiographic progression, and is sensitive to influences such as the presence of back pain and the number of somatic symptoms.¹¹

Advances in health measurement have led to the application of Rasch Item Response Theory (IRT) analysis (or Rasch analysis)¹² to assess instruments measuring health status and quality of life,¹³⁻²⁶ including the HAQ²¹ and

Table 1 WOMAC Scores in RA, OA and fibromyalgia

Variable	OA (n=655)	RA (n=1013)	Fibromyalgia (n=537)
Demographics and severity			
Age in years (SD)	67.8 (11.5)	58.8 (14.8)	55.4 (11.8)
Sex (% female)	76.5	76.1	93.9
Disease duration in years (SD)	16.9 (11.3)	9.2 (10.0)	16.5 (10.1)
WOMAC function (range 0–170) (SD)	65.1 (40.9)	53.0 (39.1)	73.8 (41.6)
WOMAC pain (range 0–50) (SD)	18.6 (11.8)	14.9 (11.4)	22.8 (12.1)
HAQ disability (range 0–3) (SD)	1.0 (0.70)	1.0 (0.75)	1.1 (0.7)
VAS pain (range 0–3) (SD)	1.3 (0.79)	1.1 (0.76)	1.7 (0.8)
Patient global severity (range 0–100) (SD)	38.7 (24.4)	35.8 (24.0)	51.7 (25.6)
Pearson correlations			
WOMAC function: correlation with HAQ	0.78	0.78	0.81
WOMAC pain: correlation with VAS pain	0.73	0.71	0.66

Except for sex, values are mean and standard deviation.

Table 2 Average WOMAC pain item calibrations, SE, and INFIT statistic ordered by calibration

WOMAC pain item	Number	Average item calibration (logits)	SE	INFIT statistic (MNSQ)
<i>Osteoarthritis</i>				
Sitting pain	655	0.39	0.03	0.88
Night pain	655	0.28	0.03	1.20
Pain walking	655	0.10	0.03	0.87
Standing pain	655	-0.06	0.03	0.96
Pain on stairs	655	-0.71	0.03	1.12
<i>Rheumatoid arthritis</i>				
Sitting pain	1013	0.25	0.02	0.94
Night pain	1013	0.11	0.02	1.16
Pain walking	1013	0.02	0.02	1.05
Standing pain	1013	0.02	0.02	0.83
Pain on stairs	1013	-0.40	0.02	1.10
<i>Fibromyalgia</i>				
Pain walking	537	0.32	0.03	0.78
Sitting pain	537	0.08	0.03	0.95
Night pain	537	0.07	0.03	1.30
Standing pain	537	0.00	0.03	0.82
Pain on stairs	537	-0.47	0.03	1.24

Negative calibrations indicate more difficult items. The more positive the score the easier the item. INFIT statistics >1.2–1.3 indicate that the item does not contribute to the underlying construct. INFIT (MNSQ) values of <0.7–0.8 indicate items that are muted.

SF-36.²⁵ Although an instrument may measure several health concepts with different subscales, a good subscale should be unidimensional, not have floor or ceiling effects, and have adequate spread along a single dimension. Individual scale items that do not fit the unidimensional model may be measuring a different concept than what was anticipated (non-

unidimensionality), may be misunderstood by respondents, or may not apply to all persons under study. An additional problem with measurement scales is that some of the items may be redundant, thereby contributing no additional information.

Studies of the SF-36 physical function scale (SF-10) have shown elements that do not have adequate fit to a unidimensional model.^{21–25} Recent Rasch IRT studies by Tennant *et al* showed that the HAQ has impaired construct validity in OA, and in rheumatoid arthritis (RA) it had ceiling effects and inadequate spread.²¹ This study examined whether similar problems might exist with the WOMAC. In addition, we have used the WOMAC in RA and fibromyalgia. Therefore we also examined whether the Rasch analysis yielded similar results across the three disorders. If results were found to be similar they would offer additional support for the use of the WOMAC in other illnesses besides OA, and they would also confirm the results of Rasch analyses in OA.

Methods

Data for this study were obtained from patients participating in a long term outcome study of OA, RA and fibromyalgia. Most patients were attending the Arthritis Center in Wichita, Kansas, an outpatient rheumatology clinic and research centre where longitudinal data on OA, RA and fibromyalgia patients had been collected since 1974. As part of the data collection, questionnaires at six month intervals were mailed to patients who chose to participate in the longitudinal study. The characteristics of this data bank and the methods of data collection have been described previously.^{27–28} The current report relied on mailings sent between July 1996 and January 1998 when the WOMAC questionnaire was added to the assessment package and included 2205 patients, 1013 with RA, 655 with OA, and 537 with fibromyalgia. Among the RA patients, 447 of the 1013 were members of a US inception cohort of RA who were recruited during the study period from the practices of rheumatologists, and who had a disease duration of less than one year when first seen by their rheumatologists. Of the 655 OA patients, 348 were recruited during the study period by media and mailed advertising for participation in an OA outcome project. Seventy five of the 537 fibromyalgia patients were from centres other than Wichita who had participated in previous fibromyalgia outcome studies.²⁹ When more than one WOMAC assessment was available, we made use of the most recent assessment. Thus we are performing cross sectional not longitudinal analyses in this study.

Patients with RA and fibromyalgia satisfied published diagnostic/classification criteria.^{30–31} Patients with OA had definite radiographic abnormality and knee pain, and clinically had OA. Although most satisfied published criteria for OA,^{32–33} it was the purpose of this project to identify mild cases so that minimal entry criteria for this study included a clinical diagnosis of OA, definite osteophytes, and characteristic knee pain.

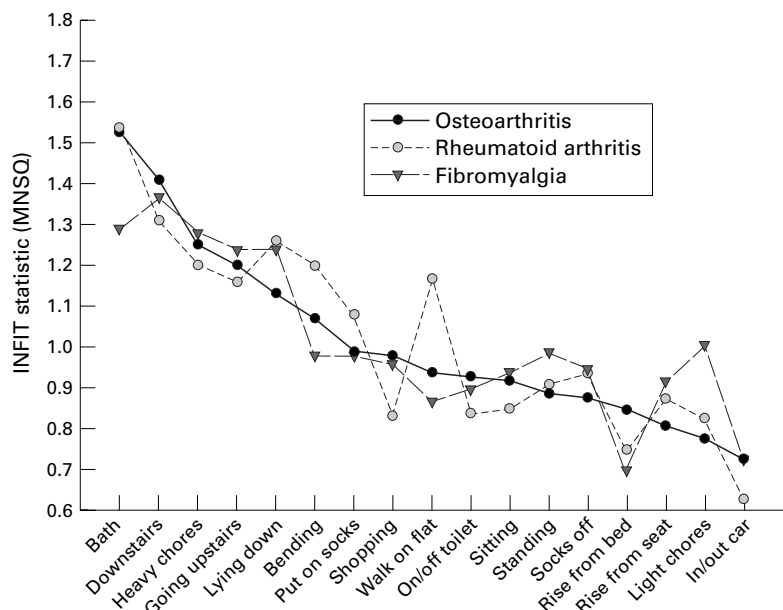


Figure 2 INFIT (MNSQ) statistics for WOMAC functional scale items. INFIT statistics > 1.2–1.3 indicate that the item does not contribute to the underlying construct. INFIT (MNSQ) values of < 0.7–0.8 indicate items that are muted.

Table 3 Average WOMAC functional item calibrations, SE, and INFIT statistic ordered by misfitting

WOMAC function: difficulty with	Number	Average item calibration (logits)	SE	INFIT statistic (MNSQ)
<i>Osteoarthritis</i>				
Getting in and out of bath	655	-0.19	0.02	1.53
Going down stairs	655	-0.27	0.02	1.41
Heavy chores	655	-0.69	0.03	1.25
Going upstairs	655	-0.47	0.02	1.20
Lying down	655	0.50	0.03	1.13
Bending	655	-0.24	0.02	1.07
Putting on socks	655	0.05	0.02	0.99
Shopping	655	-0.16	0.02	0.98
Walking on flat ground	655	0.22	0.03	0.94
Getting on and off the toilet	655	0.30	0.03	0.93
Sitting	655	0.51	0.03	0.92
Standing	655	0.04	0.02	0.89
Taking off socks	655	0.21	0.03	0.88
Arising from bed	655	-0.19	0.02	0.85
Arising from sitting	655	0.11	0.02	0.81
Light chores	655	0.33	0.03	0.78
Getting in and out of a car	655	-0.07	0.02	0.73
<i>Rheumatoid arthritis</i>				
Getting in and out of bath	1013	-0.30	0.02	1.54
Going down stairs	1013	-0.20	0.02	1.31
Lying down	1013	0.46	0.02	1.26
Bending	1013	-0.20	0.02	1.20
Heavy chores	1013	-0.81	0.02	1.20
Walking on flat ground	1013	0.17	0.02	1.17
Going up stairs	1013	-0.25	0.02	1.16
Putting on socks	1013	-0.01	0.02	1.08
Taking off socks	1013	0.11	0.02	0.94
Standing	1013	0.11	0.02	0.91
Arising from sitting	1013	0.06	0.02	0.88
Sitting	1013	0.48	0.02	0.85
Getting on and off toilet	1013	0.29	0.02	0.84
Shopping	1013	-0.10	0.02	0.83
Light chores	1013	0.22	0.02	0.83
Arising from bed	1013	-0.06	0.02	0.75
Getting in and out of a car	1013	0.02	0.02	0.63
<i>Fibromyalgia</i>				
Going downstairs	537	0.02	0.03	1.37
Getting in and out of bath	537	-0.15	0.03	1.29
Heavy chores	537	-1.04	0.03	1.28
Going upstairs	537	-0.43	0.03	1.24
Lying down	537	0.28	0.03	1.24
Light chores	537	0.25	0.03	1.01
Standing	537	0.08	0.03	0.99
Putting on socks	537	0.20	0.03	0.98
Bending	537	-0.29	0.03	0.98
Shopping	537	-0.16	0.03	0.96
Taking off socks	537	0.31	0.03	0.95
Sitting	537	0.31	0.03	0.94
Arising from sitting	537	-0.05	0.03	0.92
Getting on and off toilet	537	0.45	0.03	0.90
Walking on flat ground	537	0.29	0.03	0.87
Getting in and out of a car	537	0.02	0.03	0.73
Arise from bed	537	-0.09	0.03	0.70

Negative calibrations indicate more difficult items. The more positive the score the easier the item. INFIT statistics >1.2–1.3 indicate that the item does not contribute to the underlying construct. INFIT (MNSQ) values of <0.7–0.8 indicate items that are muted.

All patients completed the WOMAC. The multidimensional Western Ontario MacMaster osteoarthritis index assesses pain, stiffness, and physical function activities related to OA of the hip or knee.^{1 2 6 7 34} In this study the WOMAC was used in its visual analogue scale (VAS) format (version VA3.0) but was converted to an 11 step 0–10 scale for Rasch analyses. This method of using a VAS scale in Rasch analysis has been previously described.²⁵ In these analyses we studied the 17 item function scale and the 5 item pain scale, but we did not analyse the WOMAC stiffness scale because it had only two items. This version of the WOMAC questionnaire does refer specifically to any joint.

STATISTICS

Data were analysed use Stata Version 5.0³⁵ for general descriptive statistics. For Rasch analyses we used Winsteps (Bigsteps) Version 2.8.³⁶

In the Bigsteps Rasch analysis, patients and item scores are used to “calibrate” items on a logit scale where the midpoint of the scale is 0. Items at one end of the scale are “easier” and items at the other end are more “difficult.” In the current analyses, items with a negative (–) calibration are more difficult. Individual items that are at least 0.15 logits apart represent individual strata.³⁷ It is generally desirable that this separation distance between items be 0.15 logits or more. Otherwise one item is not distinctly separate from the next. Another important characteristic of a good instrument is that it has a good overall separation (expressed in logits). The greater the separation the more distinct strata are identified. But the separation between individual items should also not be too far or spaces will occur between the individual items. When the data fit the Rasch model the information weighted fit statistic (INFIT) will be between 0.7–0.8 to 1.2–1.3 using the mean square INFIT (MNSQ) statistic. When the data fit well they indicate that the subscale items contribute to a single underlying construct (unidimensionality). When the INFIT statistics are applied to the individual items, INFIT statistics > 1.2–1.3 indicate that the item does not contribute to the underlying construct or is “noisy”. An INFIT statistic of 1.3 means that there is 30% more noise than expected. Generally the higher number is taken as the upper limit of allowable “noise”. INFIT (MNSQ) values of < 0.7–0.8 indicate items that are muted. This may occur when there are several items that are similar or highly correlated, or when one item is dependent on another. Generally the lower number (0.7) is taken as the lower limit for INFIT. “A mean-square of 0.7 indicates 43% more ambiguity in the inferred measure than modeled.”³⁸ The choice of limiting values is in part a function of the purpose for which the scale will be used. Wright and Linacre indicate that reasonable fit statistics are between 0.7–1.3 for “run of mill” tests and that 0.8–1.2 are reasonable for “high stakes” tests.³⁸ Thus the scale that perfectly fits the Rasch model is unidimensional, has adequate separation so that there are sufficient strata, has items that are not calibrated too far apart, and has individual items that all contribute to the underlying construct.

Results

Figure 1 depicts the distribution of WOMAC function scores for the 2205 study participants. As shown in table 1, WOMAC function and pain were most abnormal in the 537 fibromyalgia patients and least abnormal in 1013 RA patients. The 655 patients with knee and hip OA held a middle position.

WOMAC PAIN SCALES

Rasch analyses were performed on each patient group separately. For OA patients the overall INFIT and separation statistic for pain was 1.01 (table 2) with an item separation of 13.26. Similar values were 1.02, 9.11 for RA and 1.02, 7.97 for fibromyalgia. More positive scores for average item calibration indicate easier categories, and more negative scores indicate more

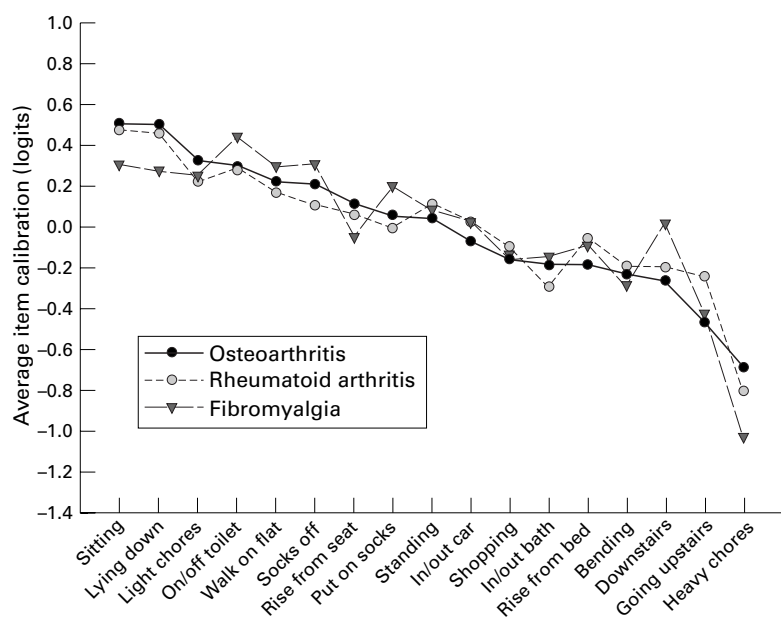


Figure 3 Average calibration in logits for WOMAC functional scale items. Negative calibrations indicate more difficult items. The more positive the score the easier the item.

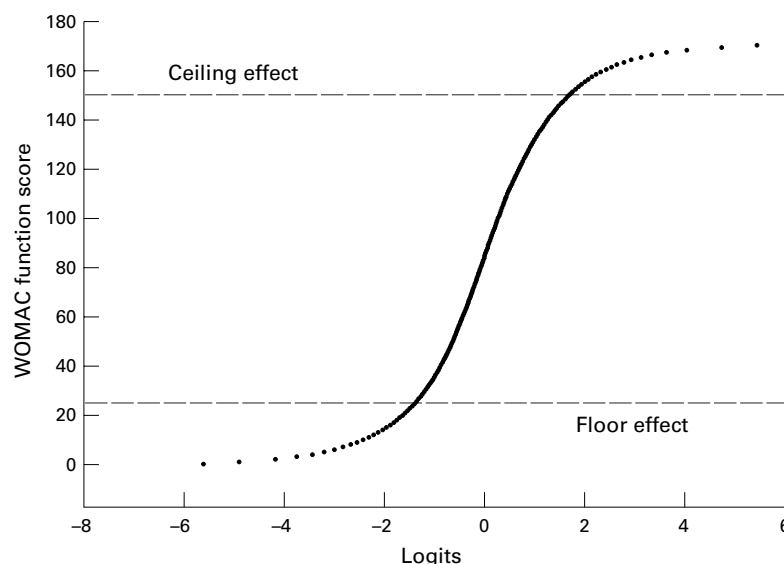


Figure 4 Plot of severity in logits versus WOMAC functional score for OA. Curves are similar in RA and fibromyalgia. The WOMAC function score is linear over the range of 25–150 (equivalent range on 0–10 scale is 1.5–8.8). Less than two per cent of patients have scores greater than 150. Twenty one per cent have scores less than 25 (see fig 1).

difficult categories. Table 2 shows that, for OA, individual inter-item differences of at least 0.15 logits were generally identified, indicating satisfactory item separation. Among RA patients, however, items about night pain, pain with walking, and pain standing were not as well separated, and this was true in fibromyalgia patients as well. For fibromyalgia patients the easiest category was walking, in contradistinction to the other groups where sitting and night pain were the easiest categories—fitting the clinical complaints that are often heard. As with RA patients, middle variables had calibrations similar to fibromyalgia. These data suggest that for pain scores the WOMAC performs appropriately in terms of INFIT and separation statistics for all patient groups, but with fewer distinct strata in the RA and fibromyalgia groups as indicated by inter-item differences of less than 0.15 logits.

WOMAC FUNCTION SCALES

As with the pain scale, general INFIT and separation statistics for function were quite satisfactory. INFIT and separation values for OA, RA and fibromyalgia were 1.02 and 12.21; 1.01 and 13.58; and 1.02 and 11.07, respectively.

The individual items of the 17 item function scale were examined to understand their appropriateness as part of a unidimensional scale. Table 3 and figure 2 present these items ordered by their INFIT statistics. As indicated in the methods section, INFIT (MNSQ) values greater than 1.2–1.3 indicate items that have unexpected values or items that may tap additional dimensions. Among OA patients, “getting in and out of the bath”, “going down stairs”, and performing “heavy chores” have scores greater than 1.20. These items also poorly fit the unidimensional model among patients with RA and fibromyalgia, although RA patients also had other poorly fitting items. These items and their relation to diagnostic group membership can be seen clearly in figure 2.

Data in table 3 and figure 2 also indicate items that may be redundant or have dependencies. These items were identified by INFIT statistics of 0.8–0.7 or less. Doing light chores and getting in and out of a car were among those items for all groups, and rising from a bed were such items for the RA and OA groups. When the INFIT criterion is set at 0.7, then only “getting in and out of a car” in the RA group is misfitting.

Table 3 and figure 3 display the item calibrations for the WOMAC function items for the three groups. Going upstairs and performing heavy chores were clearly the most difficult items. The figure demonstrates the general similarity of results among the groups as well as the differences between individual items.

Graphic examination of item calibrations (fig 4) indicated a large range (25–150) in which the WOMAC function score was linear in OA patients. The equivalent range on a 0–10 scale is 1.5–8.8. There is no substantial ceiling effect, for only two per cent of patients have scores greater than 150. By contrast, 21 per cent of patients had scores less than 25, reflecting the larger number of patients with very mild disease. Although INFIT statistics for these lower ranges (steps) indicate appropriate fit, as with scores over 150 there is also greater distance (or severity) between each step of WOMAC function in the ranges below 25 and above 150 than in the range of 25 to 150. Similar results (not shown) were obtained in RA and fibromyalgia patients.

Discussion

The data presented here indicate that the WOMAC generally performs well in OA, RA and fibromyalgia. The overall fit statistics, scaling and separation are quite satisfactory; and results obtained in the three disorders are quite similar. A few individual items, however, do not have adequate fit statistics. For items with high INFIT statistics, the item may be addressing a different process or content—that is, an item in

another dimension. Getting in and out of the bath may require arm strength or general well being; and the item might measure cardiopulmonary problems or weakness, qualities that differ from lower extremity problems. Performing heavy chores or going upstairs might have the same trouble. Going downstairs could tap into problems related to balance or other neurological conditions. Another, and more likely explanation, is that tub baths are not common, showers being used, so that the experience of patients is not uniform. Similarly, problems with stairs may reflect the limited use of stairs by many patients. These items were troublesome across all of the three disorders so that it is likely that they represent true psychometric problems for the scales.

At the other end of the spectrum, where the INFIT statistic is low, redundancies can cause poor fitting. This is true when one item answers another or is strongly correlated with it; or if there are similar items. Getting in and out of a car, doing light chores and rising from a bed were such items for the study populations.

The overall fit statistic was satisfactory in RA and fibromyalgia, although the overall separation for WOMAC pain was greater in OA than the other conditions, indicating that pain assessment was better in OA than RA and fibromyalgia.

A case could be made to reformulate the WOMAC by eliminating redundant items and eliminating items such getting in and out of the bath, going downstairs, and possibly doing heavy chores. We have recently shown that the presence of low back pain is highly correlated to WOMAC scores, as are general medical complaints.¹¹ It is possible that some of the noise that is present in the WOMAC assessment of OA comes through such mechanisms and are more expressed in items with high INFIT scores.

As we expected, we found evidence of non-linearity (ceiling and floor effects), as shown in figure 4. This non-linear effect was unimportant as a ceiling effect because only two per cent of patients had values above the ceiling. However, 21 per cent had values in the range of the floor effect. The non-linear effects have two practical considerations. Firstly, it can generally be assumed that a WOMAC score that is, for example, two or three times greater than another score and that falls within the zone of linearity, approximately is measuring severity that is two or three times greater. But no such assumption can be made for scores above the ceiling or below the floor level. Secondly, provided there are no missing data that require interpolation, ordinary statistical methods are appropriate to analyse the data when a monotonic relation between the scale and other variables is expected, as is most often the case. One way out of these problems is to apply Rasch transformations to the WOMAC scales results, thereby converting the observed scale into a linear measure. The advantage of doing this is that all of the values are then interval. Clinically, then, you can compare patients on a linear measuring scale. The benefit of doing this must be weighed against the difficulty it

imposes, a difficulty that is not inconsiderable. For use in clinical trials in which patients would be expected to have scores within the linear range, floor and ceiling effects would not be expected to be a problem, but this might not be true in observational studies of populations where many patients may have low WOMAC scores.

While high INFIT scores may have addressed elements that are not unidimensional, we do not believe this should be interpreted as a central critique of the WOMAC. Not only was the overall fit, scaling and separation of the WOMAC quite good across all of the conditions studied numerous studies have shown that the WOMAC is a sensitive and effective instrument in OA. To test whether elimination of the non-fitting items improves the usefulness of the WOMAC is relatively easy to do, by analysing a clinical trial or similar study with and without the non-fitting items. But given the general overall good fit, it is not likely that removing such items would result in more than minimal differences. In addition, it will be difficult to remove traces of multidimensionality while keeping the central constructs of progressive lower body musculoskeletal abnormality intact; and it is possible that a "purer", still more unidimensional instrument would be less useful in clinical trials and epidemiological studies by restricting the range of the scale.

There are a number of items that require comment. The WOMAC was developed on samples that included hip OA and knee OA patients. This study only uses knee OA patients. Therefore our conclusions about the WOMAC in OA only refer to knee OA. It is also true that the WOMAC was not validated nor designed for use in RA or fibromyalgia. This study deals with the psychometric properties of the instrument in these disorders. Additional studies will be required to determine if the WOMAC is useful or valid in these conditions. Even so, the WOMAC taps into dimensions that are key to fibromyalgia and are not fully evaluated by any other current instrument, and it may provide additional information (for research purposes) regarding lower body function in RA.

Funding: supported in part by grants (AR43584) from the National Institutes of Health and Merck and Co, Inc.

- 1 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
- 2 Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. *J Rheumatol* 1997;24:2413-15.
- 3 Bombardier C, Melfi CA, Paul J, Green R, Hawker G, Wright J, *et al.* Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care* 1995;33 (suppl 4):AS131-44.
- 4 Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol* 1995;22:1193-6.
- 5 Barr S, Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF, *et al.* A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC osteoarthritis index. *J Rheumatol* 1994;21:2106-12.

- 6 Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol* 1991;10:138-43.
- 7 Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF, Kraag GR, *et al*. A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. *J Rheumatol* 1993;20:999-1004.
- 8 Bellamy N, Kean WF, Buchanan WW, Gercz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC Osteoarthritis Index. *J Rheumatol* 1992;19:153-9.
- 9 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of the WOMAC: a health status instrument for measuring clinically-important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *Journal of Orthopedic Rheumatology* 1988;1:95-108.
- 10 Bellamy N, Buchanan WW. Outcome measurement in osteoarthritis clinical trials: the case for standardization. *Clin Rheumatol* 1984;3:293-303.
- 11 Wolfe F. Determinants of WOMAC function, pain, and stiffness scores: evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis and other rheumatic disorders. *Rheumatology* 1999;38:355-61.
- 12 Andrich D. *Rasch models for measurement*. Newbury Park: Sage Publications.
- 13 Granger CV, Deutsch A, Linn RT. Rasch analysis of the Functional Independence Measure (FIM) Mastery Test. *Arch Phys Med Rehabil* 1998;79:52-7.
- 14 Baker JG, Granger CV, Fiedler RC. A brief outpatient functional assessment measure: validity using Rasch measures. *Am J Phys Med Rehabil* 1997;76:8-13.
- 15 Mchorney CA, Haley SM, Ware JEJ. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *J Clin Epidemiol* 1997;50:451-61.
- 16 Mchorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. *Ann Intern Med* 1997;127:743-50.
- 17 Prieto L, Alonso J, Ferrer M, Anto JM. Are results of the SF-36 health survey and the Nottingham Health Profile similar? A comparison in COPD patients. *Quality of Life in COPD Study Group*. *J Clin Epidemiol* 1997;50:463-73.
- 18 Tesio L, Granger CV, Fiedler RC. A unidimensional pain/disability measure for low-back pain syndromes. *Pain* 1997;69:269-78.
- 19 Grimby G, Andren E, Holmgren E, Wright B, Linacre JM, Sundh V. Structure of a combination of Functional Independence Measure and Instrumental Activity Measure items in community-living persons: a study of individuals with cerebral palsy and spina bifida. *Arch Phys Med Rehabil* 1996;77:1109-14.
- 20 Nordenskiöld U, Grimby G, Hedberg M, Wright B, Linacre JM. The structure of an instrument for assessing the effects of assistive devices and altered working methods in women with rheumatoid arthritis. *Arthritis Care Res* 1996;9:358-67.
- 21 Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *J Clin Epidemiol* 1996;49:711-17.
- 22 Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA. Are we making the most of the Stanford Health Assessment Questionnaire? *Br J Rheumatol* 1996;35:574-8.
- 23 Chang WC, Chan C. Rasch analysis for outcomes measures: some methodological considerations. *Arch Phys Med Rehabil* 1995;76:934-9.
- 24 Granger CV, Ottenbacher KJ, Baker JG, Sehgal A. Reliability of a brief outpatient functional outcome assessment measure. *Am J Phys Med Rehabil* 1995;74:469-75.
- 25 Thomee R, Grimby G, Wright BD, Linacre JM. Rasch analysis of visual analog scale measurements before and after treatment of patellofemoral pain syndrome in women. *Scand J Rehabil Med* 1995;27:145-51.
- 26 Haley SM, Mchorney CA, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 1994;47:671-84.
- 27 Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: A 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. *Arthritis Rheum* 1998;41:1072-82.
- 28 Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: Work disability: A prospective 18 year study of 823 patients. *J Rheumatol* 1998;25:2108-17.
- 29 Wolfe F, Anderson J, Harkness D, Bennett RM, Caro XJ, Goldenberg DL, *et al*. Health status and disease severity in fibromyalgia: results of a six-center longitudinal study. *Arthritis Rheum* 1997;40:1571-9.
- 30 Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
- 31 Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, *et al*. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia: Report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990;33:160-72.
- 32 Altman RD, Alarcón GS, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology Criteria for the Classification and Reporting of Osteoarthritis of the Hip. *Arthritis Rheum* 1991;34:505-14.
- 33 Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. *Arthritis Rheum* 1986;29:1039-49.
- 34 Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. *Semin Arthritis Rheum* 1989;18:14-17.
- 35 Stata Corporation. *Stata Statistical Software: Release 5.0*. College Station, TX: Stata Corporation, 1997.
- 36 Linacre JM, Wright BD. *A user's guide to BIGSTEPS: Rasch model computer program*. [Computer program]. Version 2.8. Chicago: MESA Press, 1997.
- 37 Silverstein F, Kilgore DJ, Harman JG, Harvey T. Applying psychometric criteria to functional assessment in medical rehabilitation, I: exploring unidimensionality. *Arch Phys Med Rehabil* 1991;72:631-7.
- 38 Wright BD. Reasonable mean-square fit values. In: Wright BD, Linacre JM, eds. *Rasch measurement transactions*. Part 2. Chicago: MESA, 1996:370.